International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 159-164
Published Online March 2021 in IJEAST (http://www.ijeast.com)

LITERATURE SURVEY ON CAR AND PEDESTRIAN TRACKING SYSTEM

B.W. Balkhande
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering,
Navi Mumbai, Maharashtra, India.

Rohit Chaurasia
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering,
Navi Mumbai, Maharashtra, India

Pallavi Dhumal
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering,
Navi Mumbai, Maharashtra, India.

Kirti Gupta
Department of Computer Engineering
Bharati Vidyapeeth College of Engineering,
Navi Mumbai, Maharashtra, India.

***Abstract*— This project focuses on Car and Pedestrian Tracking System using Python. We are using OpenCV (Computer Vision) and HAAR Cascade Classifier Algorithm (Machine Learning Xml files). Since we are tracking both car and pedestrian using computer vision, we have considered two data sets one for car and one for pedestrian with both positive and negative images. We are using the object detection algorithms i.e. HAAR for this project. Some of the possible application can be Traffic Safety, Human - Robot Interaction, Surveillance Application and Some application in vehicle detection where it aims to provide information assisting vehicle counting, Vehicle speed Measurements, Identification of traffic accidents, Traffic flow prediction, and is also being used in Tesla Industry for their Auto Drive mode.**

Keywords— **Car Detection, Pedestrian Detection, HAAR, OpenCV, Machine Learning, Artificial Intelligence.**

I. INTRODUCTION

Considering the rapid growth and development of humans many accidents is encountered every year of Car and Pedestrian. Because of this fast growing economy and rapid urbanisation there comes an increase in revolution of vehicles world - wide. This further leads to an outsized number of crashes around the world in the last few decades. If considering pedestrian, they are one of the foremost vulnerable case as compare to vehicles and pedestrian and the collisions occurred between them which further leads to fatal injuries to all the pedestrian of all the vehicle collision case and sometimes ever good amount of fatalities. The average number of accidents happening each year is increasing day by day leading to roughly around 16% of pedestrian getting injured and 19% of the fatalities recorded worldwide.

The work described in this paper is related to developing an autonomous vision system which successfully detects the pedestrian and car in order to reduce this numbers of fatalities in future and collisions between car and pedestrians in the transportation environment.

The pedestrians deaths i.e., being accounted in 2012 was 13% and made up to 3% growth of all people which has been injured in traffic crashes under all traffic fatalities. “Consistent with the NHTSA (National Highway Traffic Safety Administration) report, a pedestrian was killed every two hours and injured every seven minutes in traffic crashes. Furthermore, crashes between vehicles and pedestrians can also end in traffic congestions and economy cost. The Increase in mortality and fatality is becoming a huge source of attention around the world”[2].

While focusing on the rapid development of the world, so as to the automobile industry, autonomous car is being introduced to the world in accordance with driverless and assisted driving technologies. Automobile industry should introduce some safety functions for the autonomous cars such as road departure warning, obstacle collision, speed maintenance functions, etc. Road, vehicle and pedestrian are the key steps for understanding autonomous driving technologies.

Car and pedestrians tracking system comprise of AI and machine learning using python and OpenCV, i.e. a huge open - source library for computer vision, machine learning and image processing. This machine learning object detection algorithm is employed to detect various objects in a picture or video input provided. This detection Technology is also employed in self driving autonomous cars.

The remaining paper consists of the following sections -

- Proposed system which is given in section 2.
- Literature Survey and Various existing algorithm which is given in section 3.
- Challenges are given in section 4.
- References are given in section 5.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 159-164
Published Online March 2021 in IJEAST (http://www.ijeast.com)

II. PROPOSED SYSTEMS

A car and pedestrian tracking system are a desktop application which detects the presence of car and pedestrian during a surveillance video. The proposed application uses machine learning and artificial intelligence to perform the task. Various machine learning techniques like car recognition and pedestrian detection are often wont to find car and pedestrian within the video. Car and pedestrian tracking system are a desktop application which detects car and pedestrian after providing video as an input on which the pre trained XML files are being executed for detection.

The main goal of our project is to specialize in detecting the car and pedestrian using one among the thing detection algorithms i.e., HAAR during this case. We are considering a little data set of car and pedestrian images within the process of coaching our algorithm.

The rise in population density and accessibility around the world for cars over the past decades has led to extensive computer vision (OpenCV) research in recognition and tracking to add a safer environment to the world. In this project we'll be detecting car and pedestrian using python with machine learning and artificial intelligence.

The system architecture of car and pedestrian tracking system is given below as follows:

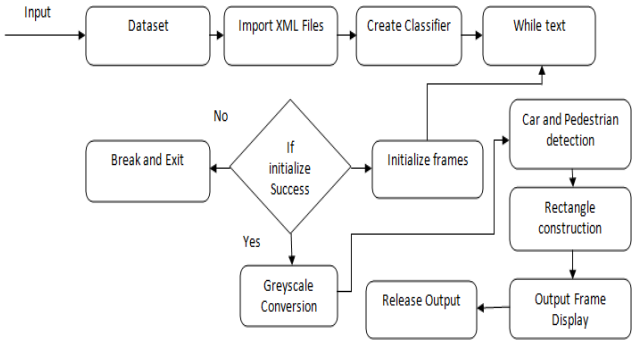


Figure: System Architecture and Work Flow Diagram of Car and Pedestrian Tracking System

Figure: System Architecture and Work Flow

HAAR Cascade Classifier Algorithm

"HAAR cascade is described as object detection algorithm in machine learning which is being mainly used to identify the objects in a picture or in video which depends on the concept of features proposed by Paul Viola and Michael Jones in their paper "Rapid Object Detection under a Boosted Cascade of Straight Forward Features" in 2001" [3].

The Proposed Algorithm has four stages:

1. HAAR Feature Selection.
2. Integral Image.
3. AdaBoost Training.
4. Cascading Classifier.

1. HAAR Feature Selection:

During this stage we select the features for our detection for an input object. This stage requires tons of images for training the classifier for detection with both positive and negative images. Then we'd like to extract feature from it, Feature are nothing but numerical information extracted from the pictures which will be use to distinguish one image from another. The various HAAR feature includes 1. Line Feature, 2. Edge Feature and 4 Rectangular Feature. By subtracting the sum of pixel under the white rectangle from the sum of pixels under the black rectangle, we obtain each HAAR feature which can be of a single value. HAAR like feature gives freedom to extract useful information from the input like edges of a video frame or of a picture, straight lines and also diagonal lines that is being used to identify the object from a picture or a video frame depending on what the input is provided, which is abbreviated as

Rectangle Feature = Sum of Pixels (Black) - Sum of Pixels (White).

"The main goal of above stage is to analyze the useful feature and decrease the processing time".

2. Integral Image:

"An Integral Image is an intermediate representation of a picture where the worth for location (x, y) on the integral image equals the sum of the pixels above and to the left (inclusive) of the (x, y) location on the first image (Viola & Jones, 2001). This intermediate representation is important because it allows for fast calculation of rectangular region"[1]. This stage resolves the matter that's being encountered in stage 1, which is in stage 1 we've to calculate the typical of a given region several time. The time complexity of those operation is O (N * N). There's numerous operations in HAAR like feature selection and reason behind that's we've to use HAAR feature with all possible sizes and site. Without Integral Image just think about what proportion of computation it would have required. Even a 24*24 window leads to at least 16,000 to 200K feature computations. Thus, we would have been required the sum of both the rectangle for every feature which eventually doubles the computation level. Hence to reduce this we use the concept of Integral Image.

3. AdaBoost Training:

It is an essential stage in HAAR Cascade Classifier algorithm, it basically apply each feature in all training frame in order to successfully find the best threshold feature that classifies the object with high accuracy as positive or

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 159-164
Published Online March 2021 in IJEAST (http://www.ijeast.com)

negative i.e. (object detected or non - detected), while the above process in motion there will be some errors or misclassification. Thus we select the feature which contains the minimum error rate that means these are the features that best classifies the object in an image or in an video frame.

After calculating this we get the final classifier which is the weighted sum of weak classifier. It is called weak because it failed to classify the object in the input. But by combining with other weak classifier together they form a strong classifier with the help of adaboost training we can detect. For example, a face with just 200 features which provides us 95% accuracy of detection so with these strong classifier one can successfully detects any object with just 6000 features instead of 200K features. This algorithm which is being used in adaboost training is same (stage wise additive multi modeling using multi class exponential class function). Decision stage is used in this algorithm which contains only one value for its confirmation of weak classifier. Performance of these stumps can be calculated as,

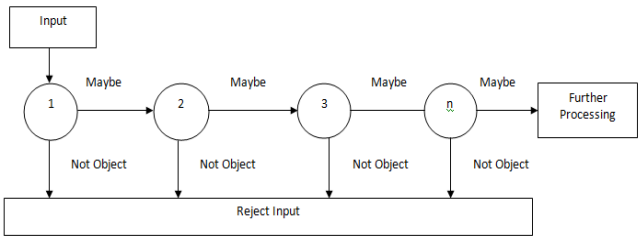
Stumps = ½*$\log_e$ ((1 - Total Errors)/Total Errors)

4. Cascading Classifier:

The process described in above its stage is quite efficient but a major issue still remains in a video frame or an image most of the image is non-object region giving equal importance to each region of the frame/image makes no sense, Thus we should mainly focus are the region of objects in a frame. Viola and Jones achieved an increased detection rate while reducing computation time using cascading classifier. The classifier is trained using adaboost and adjusting the threshold to minimize the false rate when training such model, the variables are the following:

1. The Number of Classifier stage.
2. The Number of features in each stage.
3. The threshold of each stage.

The more numbers of stages we performed the more accuracy we get.



III. LITERATURE SURVEY

[1] HOG Detector – Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) algorithms is developed by "Navneet Dalal" and "Bill Triggs" in 2005. HOG is a feature descriptor which can be used in image processing, mainly for object detection. A feature descriptor may be is in the form of a picture or a picture patch that simplifies the image by segregating useful information from it. The local object appearances within a picture are mainly described by the separation of intensity gradients or edge direction is the principle behind the histogram of oriented gradient descriptor. The x and y derivatives of a picture (Gradients) are useful because the magnitude of gradients is wide around the edges and corners, thanks to quick change in intensity and we also know that the edges and corners packs having a depth information about object shape then flat region. This is the purpose behind using histogram of direction of gradient as a feature during this descriptor.

Workflow of object detection using HOG:

"Now that we all know fundamentals of Histogram of Oriented Gradients we'll be getting into how we calculate the histograms and the way these feature vectors, which are obtained from the HOG descriptor, are employed by the classifier such a SVM to detect the concerned object"[20][22].

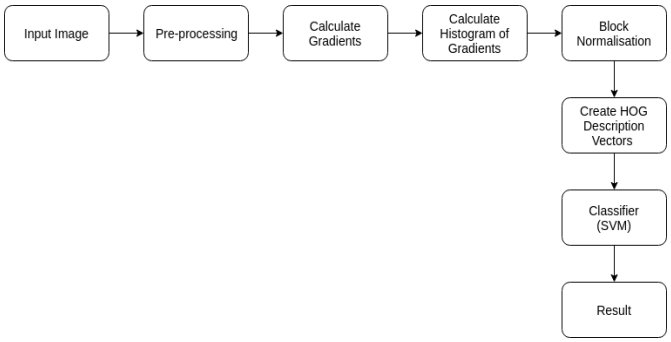


Figure: Steps for Object Detection with HOG

Advantages of HOG:
1. HOG features are extracted from the CPU or computing cluster.
2. HOG based classifiers are preferred over other classifiers as they are fast to coach and evaluate, which provides confidence level for training multiple classifier.
3. HOG based classifier achieved 79.119% accuracy rate which gives a low false positive and false negative rates.

Disadvantages of HOG:
1. The limitation of HOG is that the thing must be within a "perfect" area on the screen. Neither too closes nor too far, otherwise won't it detect the thing.
2. HOG based classifier gives a lesser accuracy in image rotations. Thus, HOG cannot be used as a good selection for classification of textures or objects which often detects for rotated image.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 159-164
Published Online March 2021 in IJEAST (http://www.ijeast.com)

[2] Local Binary Pattern (LBP) -

The algorithm is pretty straight forward and informative. The texture operator is very efficient which labels the pixel of the image by thresholding the neighborhood of every pixel and encounters the binary number as result. "Because of this discriminative power and computational simplicity, LBP texture operator has become a popular approach in several applications. The foremost important property of the LBP algorithm operator in real-world applications is its robustness to monotonic gray-scale changes caused, as an example, by illumination variations. Another important property is its computational simplicity, which makes it possible to research images in challenging real-time settings"[21].

Advantages of LBP:
1. The extent of detection of object is fast.
2. Greater accuracy level.
3. Low complexity.

Disadvantages of LBP:
1. The extent of recognition remains lacking.
2. The time needed for recognition is long enough.

[3] Single-Shot Detector (SSD)

SSD has comprised of two components mainly:
1. Backbone model
2. SSD head.

Backbone model usually may be a pre-trained image classification network as a feature extractor. this is often typically a network like ResNet trained on ImageNet from which the ultimate fully connected classification layer has been removed. Thus we have left with a deep neural network which is ready to extract semantic meaning from the input image while preserving these spatial structures of the image at a lower resolution. For ResNet34, the backbone leads to a 256 7x7 feature maps for an input image.
We'll explain what feature and have map are afterward. The SSD head is simply one or more convolutional layers added to the present backbone and therefore the outputs are interpreted because the bounding boxes and classes of objects within the spatial location of the ultimate layers activations.
Low resolution images are used for higher accuracy and speed of SSD. A growth decrease in convolutional filter for predicting object categories and bounding box locations in offsets.
High detection accuracy in SSD can be achieved by using many boxes or filters in different sizes band ratios for detection of objects. All these filters are applied to multiple feature maps from the remaining stages of network. Thus multiple scales can be formed by performing detection.

Advantages of SSD:
1. SSD may be a single-shot detector. It predicts the boundary boxes and no delegated region proposal network. A feature map in single pass refers to classes.
2. The accuracy may be enhanced single-shot detector introduces filters to predict objects, classes as well as offers to default boundary boxes.

Disadvantages of SSD:
1. The single shot detector doesn't work for smaller objects when compared to bigger objects.
2. The necessity for complex data augmentation suggests a need of an outsized knowledge to coach.

IV. CHALLENGES

Many challenges are faced so as to urge the expected outcomes. Few of the challenges faced are as follows :

Storage

A video may contain many faces. Especially during a crowded area, where many of us are present at an equivalent time. So as to form the appliance more accurate, every face within the video should be extracted in order that no face is missed by the appliance. Additionally to the present the dimensions of the videos may vary and a few video could also be large in size. Large video contain sizable amount of frames from which faces are extracted. Thanks to this an outsized number of faces are to be saved. Hence, the dimensions of the storage is large.

Faces in background

An image frame or video frame can contain different faces in various positions. Some within the front of the scene which appear clear and distinct while some within the back which can not appear as clear and distinct because the other faces. The features of the faces might not be clear. Hence it becomes difficult to spot the faces present at the rear in any given video.

Partial faces

In a particular video frame, sometimes it's going to happen that an individual doesn't enter the video frame completely. The face of the person may appear partially within the video. In such times it becomes difficult to detect the face within the video. It's going to also happen that an individual isn't facing the camera but, is facing sideways. In such cases, it becomes difficult to spot an individual to be the target person only by seeing its side face. Detecting a face leaned sideways then comparing it to the target face is additionally difficult because the orientation of the face may change the values of the eigen vector.

Similar looking people

Often it's going to happen that two people look almost like one another up to some extent. Thus, one person can easily be confused with another then are often their faces. It becomes important that similar looking people are often distinguished with one another to avoid the confusion of identifying the incorrect person because the target person.

No training set

As the image of the target person is provided by the user in real time, no training set is out there to coach the model for identifying the target person. We match the face of the target person on to the faces extracted from the videos. It might even be inconvenient to ask the user for multiple images of the target one that is to be found within the videos. Thus, it becomes difficult to spot the faces without a training set.

"The moving object may be a non rigid thing that moves over time in image sequences of a video captured by a fix or moving the camera. In video closed-circuit television the region of interest may be a person that must be detected and tracked within the video. However, this is often not a simple task to try to thanks to the various challenges and difficulties involved. These challenges occur at various different levels of object (pedestrian) detection. Video acquisition, human detection and its tracking"[13][19]. Thus occurs a plenty numbers of challenges in a video frame or in an image frame or in an image frame which are high quality frames, shadows, deformation of various objects, complex background, etc. In pedestrian detection and tracking the various challenges that occurs are the area with high crowd density tracking, occlusion and different poses.

V. REFERENCES

[1] John Jagtiani, Chris Fotache, Hessie Jones, Nick Cox, Toronto, Ontario, founded in 2016, specialities in Data Science, Machine Learning, Artificial Intelligence and community building <www.towardsdatascience.com>

[2] St. Alban-Anlage 66, Academic Open Access Publishing, Scientific Open Access Journals, and Academic Conferences, founded 1996 <www.mdpi.com>

[3] Kumar Mangalam Birla, Prof. Souvik Bhattacharyya, Prof. Sudhirkumar Barai, Birla institute of Technology and Science, Pilani<www.birla.com/studentpaper>

[4] Philomin Vasanth, Duraiswami Ramani, Davis Larry S. (13/ December /2000) Pedestrian Tracking From a Moving Vehicle, Research_Gate.

[5] Shuai Hui, Liu Qingshang, Zhang Kaihua, Yang Jing, Jiankang Deng in Wadsworth (1993) "Cascaded Regional Spatio-Temporal Feature-Routing Networks for Video Object Detection" (pp.123–135).

[6] A beginner's guide to using Python for performance computing. SciPy.org. http://scipy.github.io/old wiki/pages/PerformancePython.

[7] A. Prioletti, A. Møgelmose, P. Grisleri, M. M. Trivedi, A. Broggi and T. B. Moeslund, (Sept. 2013) "Part-Based Pedestrian Detection and Feature-Based Tracking for Driver Assistance: Real-Time, Robust Algorithms, and Evaluation, IEEE Transactions on Intelligent Transportation Systems (vol. 14, no. 3) (pp. 1346-1359) URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=& arn umber=6522156&isnumber=6587572

[8] P. Viola, M. Jones. IJCV (2004). Robust Real-Time Face Detection.

[9] S. Zhang. (2014) IEEE Transactions on Intelligent Transportation Systems.

[10] Informed HAAR-like features for Pedestrian Detection, Sigberto Alarcon Viesca, Stanford University, Stanford, CA, salarcon@stanford.edu, Brandon Garcia, Stanford University, Stanford, CA, bgarcia7@stanford.edu.

[11] OpenCV, Python, PyPI, MIT License, Olli-Pekka Heinisuo, released 2 January 2021, <https://pypi.org/project/opencv-python>

[12] Open Source, Computer Vision, by doxygen, OpenCV modules, accessed 23 April 2021, . <https://docs.opencv.org/2.4/modules/objdetect/doc/cas cade_classification.html?highlight=detectmultiscale>

[13] Pedestrian Detection and Tracking in Video Surveliance System by Ujjawala Gawande, Kamal Hajari and Yogesh Golhar, reviewed 9 December 2019, <https://www.intechopen.com/books/recent-trends-in computational-intelligence/pedestrian-detection-and tracking-in-video-surveillance-system-issues comprehensive-review-and-chall>

[14] S. Sivaraman and M. M. Trivedi (2013) "A review of recent developments in vision-based vehicle detection," 2013 IEEE Intelligent Vehicles Symposium (IV), Gold Coast, QLD, Australia (pp. 310-315) (doi: 10.1109/IVS.2013.6629487)
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arn umber=6629487&isnumber=6629437

[15] Y. Liu, B. Tian, S. Chen, F. Zhu and K. Wang (2013) "A survey of vision-based vehicle detection and tracking techniques in ITS," Proceedings of 2013 IEEE International Conference on Vehicular Electronics and Safety, Dongguan, China (pp. 72-77) (doi: 10.1109/ICVES.2013.6619606) URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6619606&isnumber=6619591

[16] Li, Zhenjiang, Wang, Kunfeng, Li, Li, Wang, Fengyu) A Review on Vision-Based Pedestrian Detection for Intelligent Vehicles, Research_Gate (13/January/2007).

[17] Wu, Juan, Peng, Bo, Huang, Zhenxiag, Xie, Jietao, Research on Computer Vision-Based Object Detection and Classification (01/January/2013)

[18] Ghogale Shweta, Bamidele Ola, Madhushree M, The International Journal of Innovative Research in Computer and Communication Engineering (issued 4 April 2021) <www.ijircce.com>

[19] Mourou Gerard, Ohsumi Yoshinori, Kroto Harold, Zimmermann Niklaus, Intechopen, from the year of 2014 – 2018 <www.intechopen.com>

[20] Joseph Coco, T.S. Lowry, Kayt Molina, Medium Member,< https://medium.com/analytics-vidhya/a-gentle-introduction-into-the-histogram-of-oriented-gradients-fdee9ed8f2aa>

[21] Dr. R. Radhika, Dr. Mohamed Abd El-Basset Matwalli, Dr. Mokhtar Beldjehem, Information Technologies <www.ijcsit.com>

[22] Chatterjee Aditya, Software Developer, Community Leader of OpenGenus Foundation, (January 2015) <https://iq.opengenus.org/object-detection-with-histogram-of-oriented-gradients-hog/>

[23] Silva Sabrina, GroundAI: A Novel Community Peer Review Platform, (November 2018) <www.groundai.com>

VI. ACKNOWLEDGEMENT

This paper was prepared under the guidance of Prof. B.W.Balkhande. The authors wish to express their sincere gratitude for providing the idea, which indeed helped us in doing a lot of Research and we came to know about many new features and technology. We would also like to thanks others faculty members who helped us to research and completion of this paper. We also extend our heartiest acknowledge to our parents for encouraging us.